Result:

**Linear regression:**

Out of sample mean square error is 1601.103, and root mean square error of linear regression is 40.01378.

**Lasso:**

图表

描述已自动生成

Lambda between -6 and -4.

Out of sample root mean square error of linear regression is 40.1186.

**Regression Tree:**

图示

描述已自动生成

full size

图表

描述已自动生成

Best size

图示

描述已自动生成

Pruning tree

The best size = 6

Out of sample root mean square error of linear regression is 43.32149

**Random forest:**

图片包含 文本

描述已自动生成

Feature importance

Out of sample root mean square error of linear regression is 39.9963

**Comparation:**

图表, 箱线图

描述已自动生成

According to the figure of root-mse comparation, random forest is the best model in this data set.

Predict result:

We used 1/5 of sample as the predict dataset, and use the best model, random forest model, to predict the value of this sample. We define the accuracy of the predict model using the formulary:

accuracy = 1 - abs(predict value – true value)/true value.

Although there are some extreme predict results, the accuracy of 3/4 of the sample are higher than 0.6, and the accuracy of half of sample are higher then 0.8. Our model has a good performance in predict dataset.

表格

描述已自动生成